

OCR: Scene text recognition

Владимир Бондаренко
Алексей Пантюхов
Александр Иванов
Пётр Марельтуев



Постановка задачи

Разработать сервис распознавания текста на объектах реального мира.

Задачи:

- Модель
- Бэкэнд
- Веб интерфейс
- Развертывание



Пайплайн обработки изображения

Изображение

Детектирование
текста

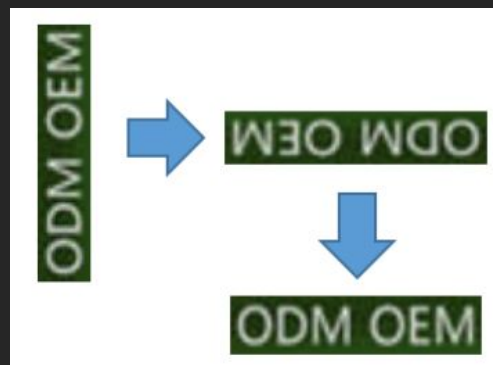
Ректификация
границ

Распознавание
текста

Текст



ResNet18

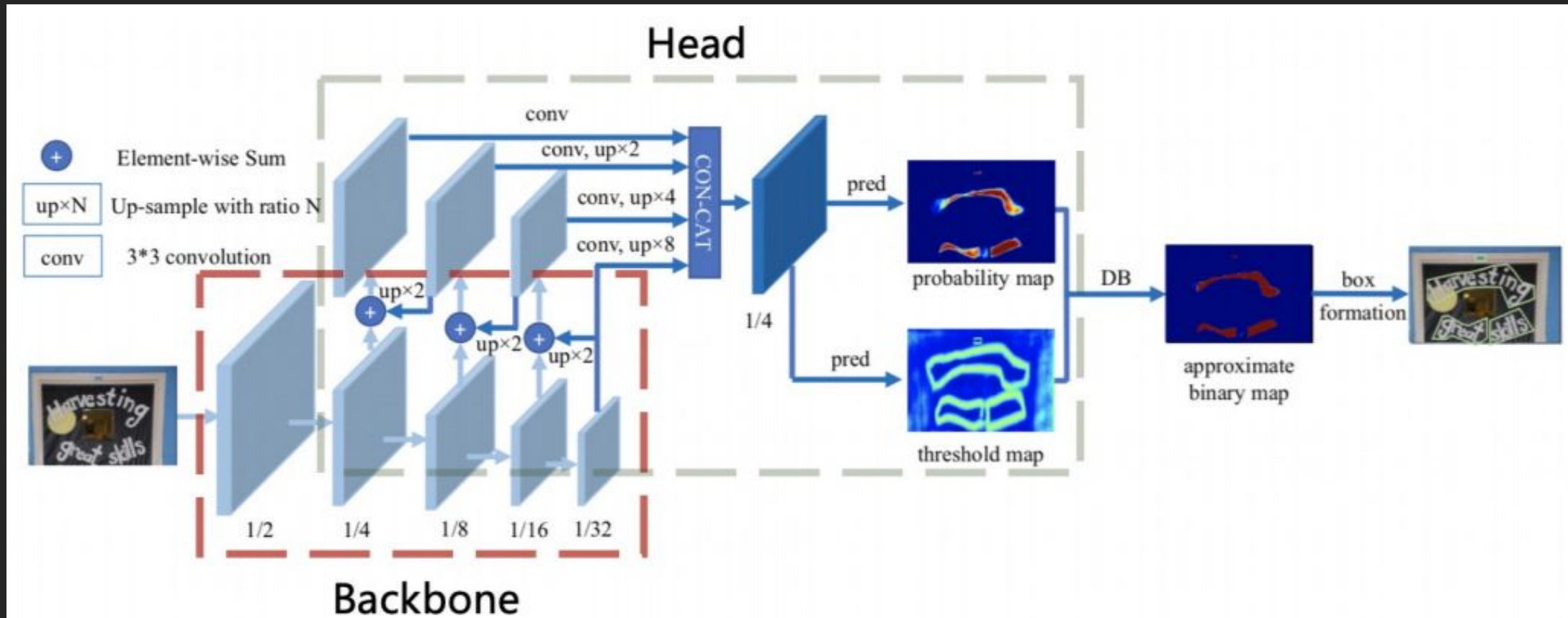


MobileNetV3



CRNN

Детектирование текста



* Backbone: ResNet18

**pred состоит из свертки 3×3 и двух деконволюционных операторов с шагом 2

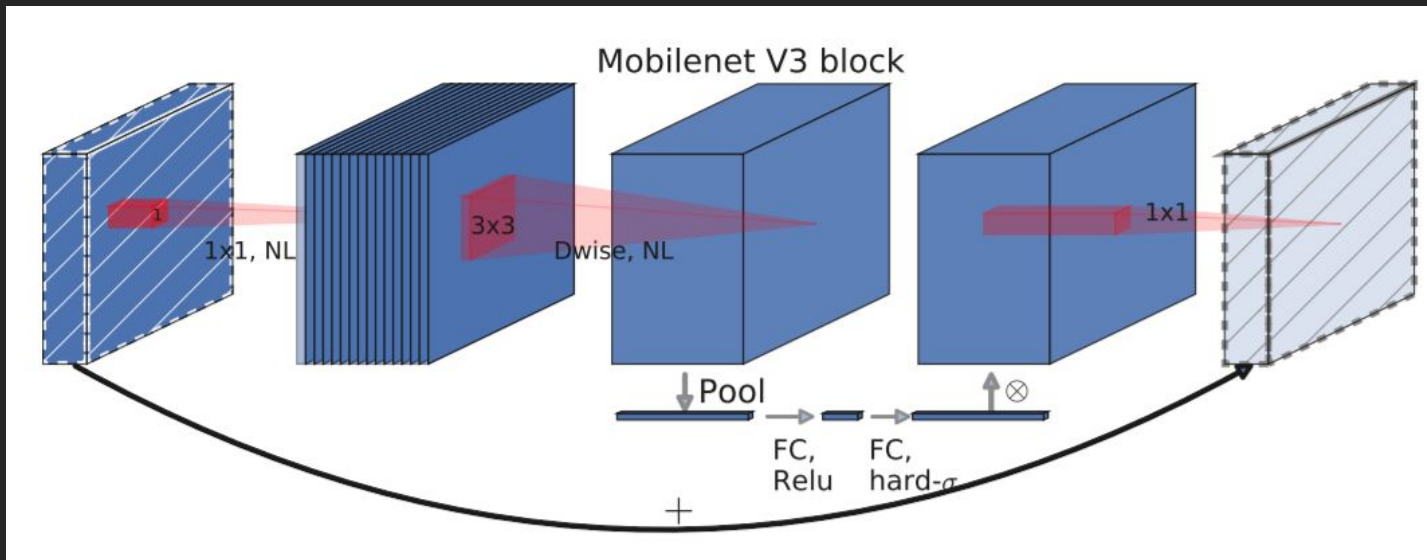
<https://arxiv.org/pdf/1911.08947>



LIFE IS FOR SHARING.

Ректификация границ

MobileNetV3

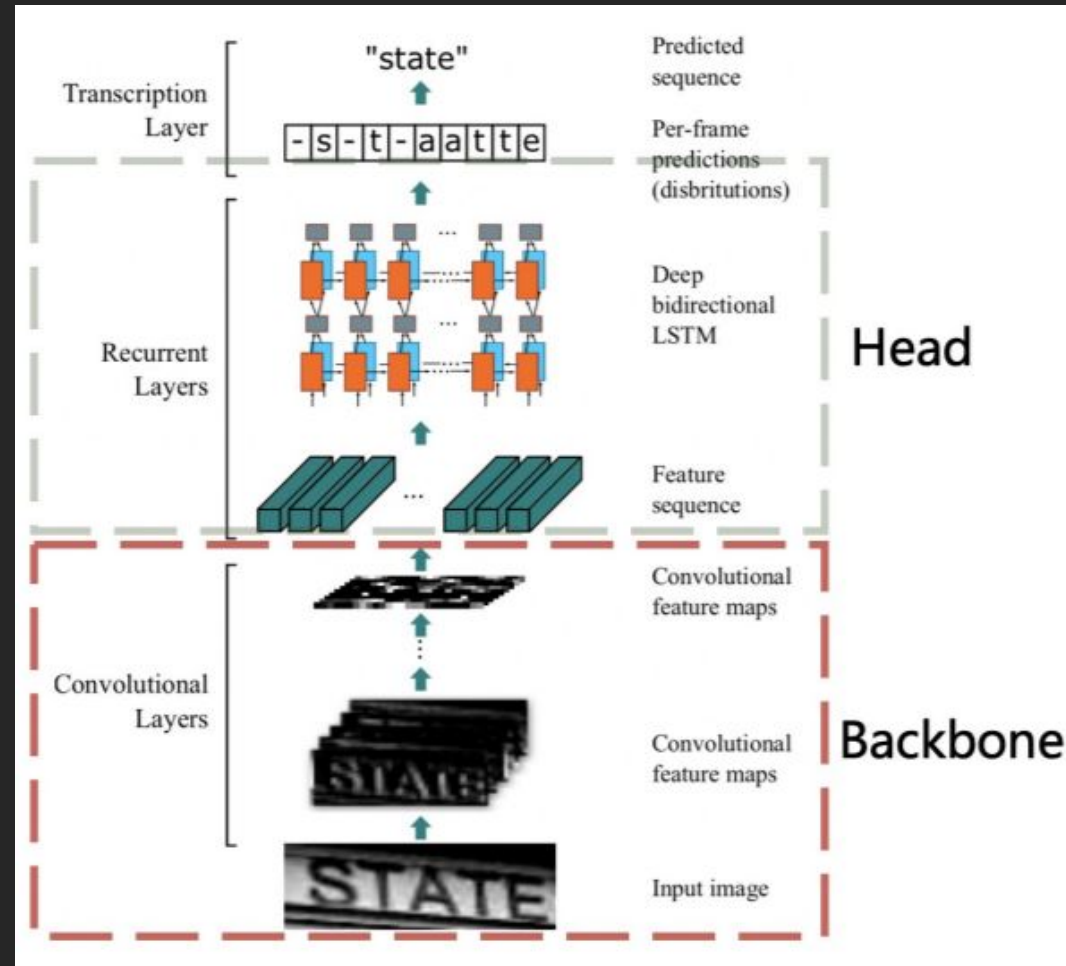


Backbone	Accuracy	Model Size (M)	Inference Time (CPU, ms)
MobileNetV3_small_x0.5	0.9494	1.34	3.22
MobileNetV3_small_x0.35	0.9403	0.85	3.21
ShuffleNetV2_x0.5	0.9017	1.72	3.41

<https://arxiv.org/abs/1905.02244>

Распознавание текста

CRNN



Метрики

Статистика по датасету

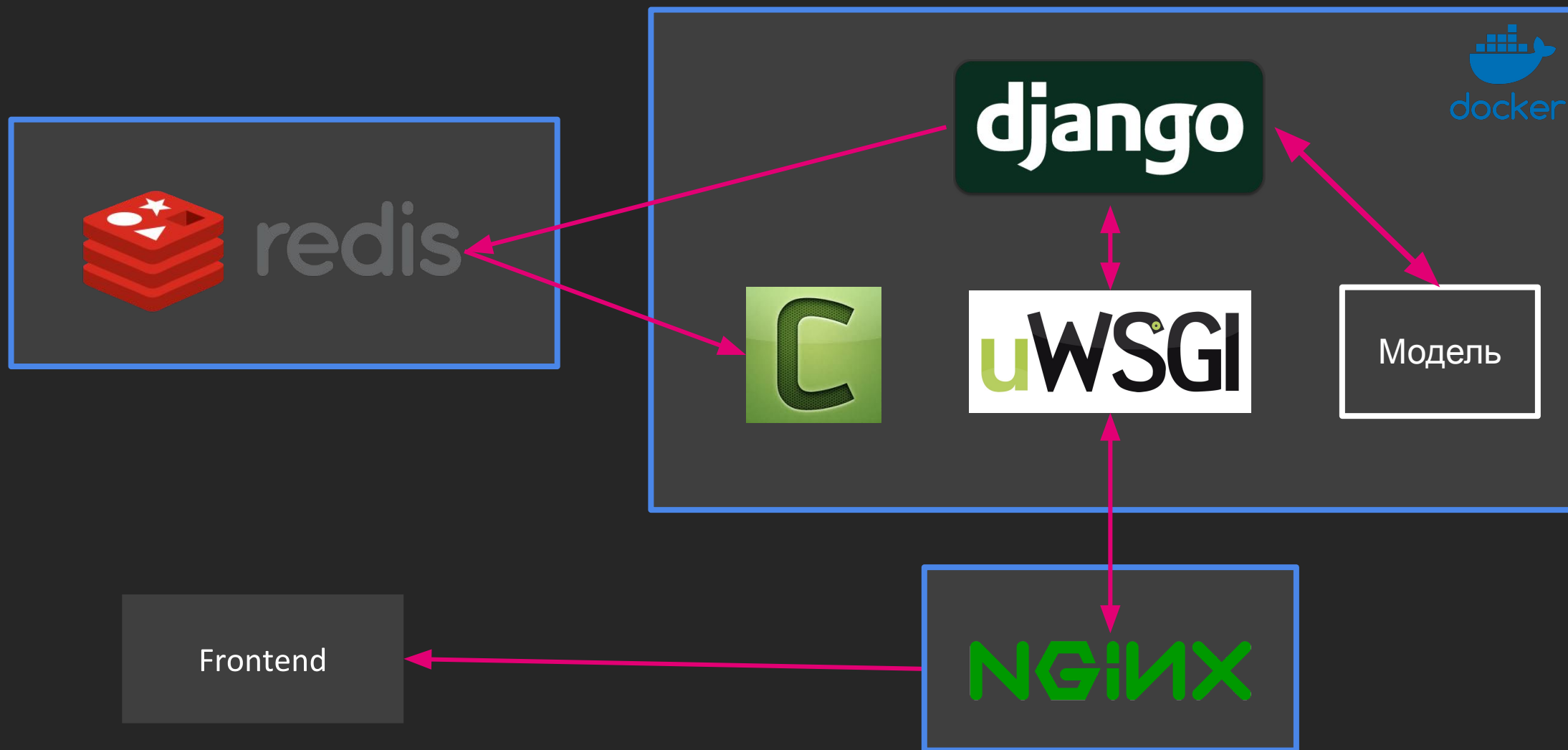
Task	Number of training data			Number of validation data
	Total	Real	Synthesis	Real
Text Detection	97K	68K	29K	500
Direction Classification	600K	100K	500K	310K
Text Recognition	17.9M	1.9M	16M	18.7K

Метрики для системы в целом

Model Type	F-score	Model Size (M)	Inference Time (ms)	
			CPU	GPU
Ultra lightweight	0.5193	8.1	421	137
Large scale	0.5414	155.1	1199	204



Архитектура системы



Ссылки

- Оригинальная статья. PP-OCR: A Practical Ultra Lightweight OCR System:
<https://arxiv.org/pdf/2009.09941>
- Репозиторий с проектом: <https://github.com/petrmareltuev/OCR>
- Real-time Scene Text Detection with Differentiable Binarization:
<https://arxiv.org/pdf/1911.08947>
- Searching for MobileNetV3: <https://arxiv.org/abs/1905.02244>
- ICDAR2017 Competition on Reading Chinese Text in the Wild (RCTW-17):
<https://arxiv.org/abs/1708.09585>

Время для демонстрации.